The 4-meter space telescope for investigating extrasolar Earth-like planets in starlight: TPF is HST2

Robert A. Brown*[a], Christopher J. Burrows†[b], Stefano Casertano*[a], Mark Clampin*[a], Dennis Ebbets‡[c], Eric B. Ford§[d], Kenneth W. Jucks#[e], N. Jeremy Kasdin**[f], Steven Kilston‡[c], Marc J. Kuchner#[e], Sara Seager††[f], Alessandro Sozzetti#‡‡[e,g], David N. Spergel§[d], Wesley A. Traub#[e], John T. Trauger§§[h], Edwin L. Turner§[d]

[a]Space Telescope Science Institute; [b]Metajiva; [c]Ball Aerospace & Technologies Corp.; [d]Princeton University Observatory; [e]Harvard-Smithsonian Center for Astrophysics; [f]Department of Mechanical and Aerospace Engineering, Princeton University; [f]Institute for Advanced Study, School of Natural Sciences; [g]Department of Physics & Astronomy, University of Pittsburgh; [h]Jet Propulsion Laboratory

ABSTRACT

Recent advances in deformable mirror technology for correcting wavefront errors and in pupil shapes and masks for coronagraphic suppression of diffracted starlight enable a powerful approach to detecting extrasolar planets in reflected (scattered) starlight at visible wavelengths. We discuss the planet-finding performance of Hubble-like telescopes using these technical advances. A telescope of aperture of at least 4 meters could accomplish the goals of the Terrestrial Planet Finder (TPF) mission. The '4mTPF' detects an Earth around a Sun at five parsecs in about one hour of integration time. It finds molecular oxygen, ozone, water vapor, the 'red edge' of chlorophyll-containing land-plant leaves, and the total atmospheric column density—all in forty hours or less. The 4mTPF has a strong science program of discovery and characterization of extrasolar planets and planetary systems, including other worlds like Earth. With other astronomical instruments sharing the focal plane, the 4mTPF could also continue and expand the general program of astronomical research of the Hubble Space Telescope.

Keywords: Deformable mirrors, coronagraphs, and extrasolar planet detection.

1. INTRODUCTION

Imaging of extrasolar planets was probably first mentioned as a NASA objective in the 1977 announcement of opportunity for the Hubble Space Telescope[1]. However, the required technology was unavailable at that time, so of course Hubble scientists and engineers did not develop the specifications and designs to achieve this goal, nor could the development contractors attempt to implement them. Indeed, it was not until the early 1980s that Elson and colleagues[2,3] provided a formalism for computing the amount of starlight that wavefront errors scatter onto the planet image, which is the dominant source of background after a coronagraph suppresses aperture-diffracted starlight. Using the Elson formalism, Brown and Burrows[4] concluded, just prior to the launch of the Hubble Space Telescope in 1990, that planet imaging by the first generation of Hubble instruments was "unfavourable". In 1997, with co-authors Traub and Trauger and other colleagues, they proposed the Coronagraph Optimized for Discovery and Exploration (CODEX) as an instrument for Hubble[5]. CODEX would correct wavefront errors to 1/2500 wave at 632.8 nm using deformable mirror (DM) technology demonstrated at the Jet Propulsion Laboratory (JPL).

*rbrown@stsci.edu; stefano@stsci.edu; clampin@stsci.edu; Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218; †chrisatmd@aol.com; Metajiva, 12320 Scenic Dr., Edmonds, WA 98026; ‡debbets@ball.com; skilston@ball.com; Ball Aerospace & Technologies Corp., P. O. Box 1062, Boulder, CO 80306; §eford@astro.princeton.edu; dns@astro.princeton.edu; elt@astro.princeton.edu; Princeton University Observatory, Peyton Hall, Ivy Lane, Princeton, NJ 08544; #kjucks@cfa.harvard.edu; mkuchner@cfa.harvard.edu; asozzett@cfa.harvard.edu; traub@cfa.harvard.edu; Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138; **jkasdin@Princeton.edu; Department of Mechanical and Aerospace Engineering, Princeton University, D207 Engineering Quadrangle, Princeton, NJ 08544. ††seager@ias.edu; Institute for Advanced Study, School of Natural Sciences, Einstein Dr., Princeton, NJ 08540. ‡‡alex@bruno.phyast.pitt.edu; University of Pittsburgh, Department of Physics & Astronomy, Pittsburgh, PA 15260 §§jtt@lyot.jpl.nasa.gov; Jet Propulsion Laboratory, MS 183-900, 4800 Oak Grove Dr., Pasadena, CA 91109

By 2002, the same DM technology proposed for CODEX/Hubble was regularly achieving 1/13000 wave performance on the JPL test bed. This advance has propelled Hubble-like telescopes to the forefront of the design contest for NASA's Terrestrial Planet Finder (TPF) mission. TPF is meant to find and study Earth-like planets around nearby stars and seek signs of life. NASA has just concluded a two-year 'architecture study' of TPF in which many of the current authors participated. As an outcome, NASA selected two architectures for further technology development: the visible-light coronagraph and the infrared nulling interferometer. In 2005, NASA plans to choose one of these two approaches for development.

This paper explains and discusses the new potential to study Earth-like extrasolar planets at visible wavelengths using a Hubble-type telescope equipped with a coronagraph and a DM-based corrective optics system. After discussing general considerations of DM performance, planet image contrast, and telescope aperture size, we connect system-level technical performance to planetary parameters and outline the science program of a particular Hubble-type TPF mission, the 4-meter version, which we call the 4mTPF. With additional instruments, this telescope could also serve as an optical-ultraviolet successor to the Hubble Space Telescope, which NASA plans to retire in 2010.

2. GENERAL CONSIDERATIONS: DM, CORONAGRAPH, CONTRAST, AND APERTURE SIZE

As illustrated in Fig. 1, a Xinetics DM with 21 x 21 actuators on 0.1 cm spacing can be commanded flat to 0.025 nm rms surface error in the range of spatial frequencies over which it has authority[6]. This result means σ_p = 0.05 nm or 1/13000 wave, where σ_p is the rms wavefront error integrated to N = 21 cycles over the pupil. This DM technology is designed to be—and successfully has been—butted for expansion. Thus, a larger DM array performing at the 1/7000-wave level (σ_p = 0.09 nm), say, can be built and operated in space, provided only that the butted interfaces are also stable at that level and thermal stability tolerances (~10 mK) are met. This paper pursues the implications of those assertions for extrasolar planet observations in visible, reflected (scattered) starlight.

For wavelength of light λ, a wavefront error of spatial wavelength Λ scatters light at angle λ/Λ in an amount proportional to the square of the error. For a continuum distribution of errors, the point-spread function (PSF) is proportional to the power spectral density (PSD) of errors. As expected for DM errors uncorrelated between actuators, the JPL test-bed DM produces a flat PSD in the corrected range of spatial frequencies. This means that the distribution of scattered light from a centered star—the point-spread function (PSF)— in the detection zone on the focal plane of a telescope of pupil diameter D using this DM technology will also be flat: $PSDPSF = 4\,\pi\,(\sigma_p/\lambda)^2\,(D/N\lambda)^2$ for a DM with 2N actuators across the pupil. (Aside from efficiency factors, the count rate for scattered starlight in a detector pixel is the product of the PSDPSF, the incident stellar flux, the area of the primary mirror, and the solid angle on the sky subtended by the pixel.)

We envision using a DM inside a coronagraph on a Hubble-type ultraviolet-optical telescope, located at a re-imaged pupil and correcting wavefront errors to 1/7000 wave rms, say, between 1 and N cycles across the pupil. The coronagraph suppresses diffracted starlight to at least the level of the PSDPSF. The minimum amount of 'Airy suppression' required—the ratio of the PSDPSF to the Airy PSF peak—is $(4/N)^2\,(\sigma_p/\lambda)^2$. This suppression applies in the annular 'detection zone', which is the range of field angles from the star between the inner-working distance (IWD) and the outer working distance (OWD), which are approximately $4\lambda/D$ and $N\lambda/D$, respectively.

The contrast ratio (Q) is the central surface brightness of the planet image divided by the surface brightness of the local scattered starlight. In the detection zone, the contrast ratio $Q = s\,C\,N^2 / 16\,(\sigma_p/\lambda)^2$, where the factor s ~ 0.5 accounts for the expected reduction in the planet image peak by low-order aberrations and the coronagraph's pupil mask, and C is the ratio of the planet flux to the star flux. C depends on the planet's geometric albedo and phase function, its radius, the viewing geometry, and the planet-star distance.

The aperture size D is important for planet detection for at least three reasons. First, because $IWD \sim D^{-1}$, the number of stars available for searching to a given minimum orbit size is proportional to D^3. Second, $Q \sim D^2$ for the same OWD. Third, the integration time $T \sim D^{-2}$ for the same Q. Next, we will closely examine the case of D = 4 meters (4mTPF).

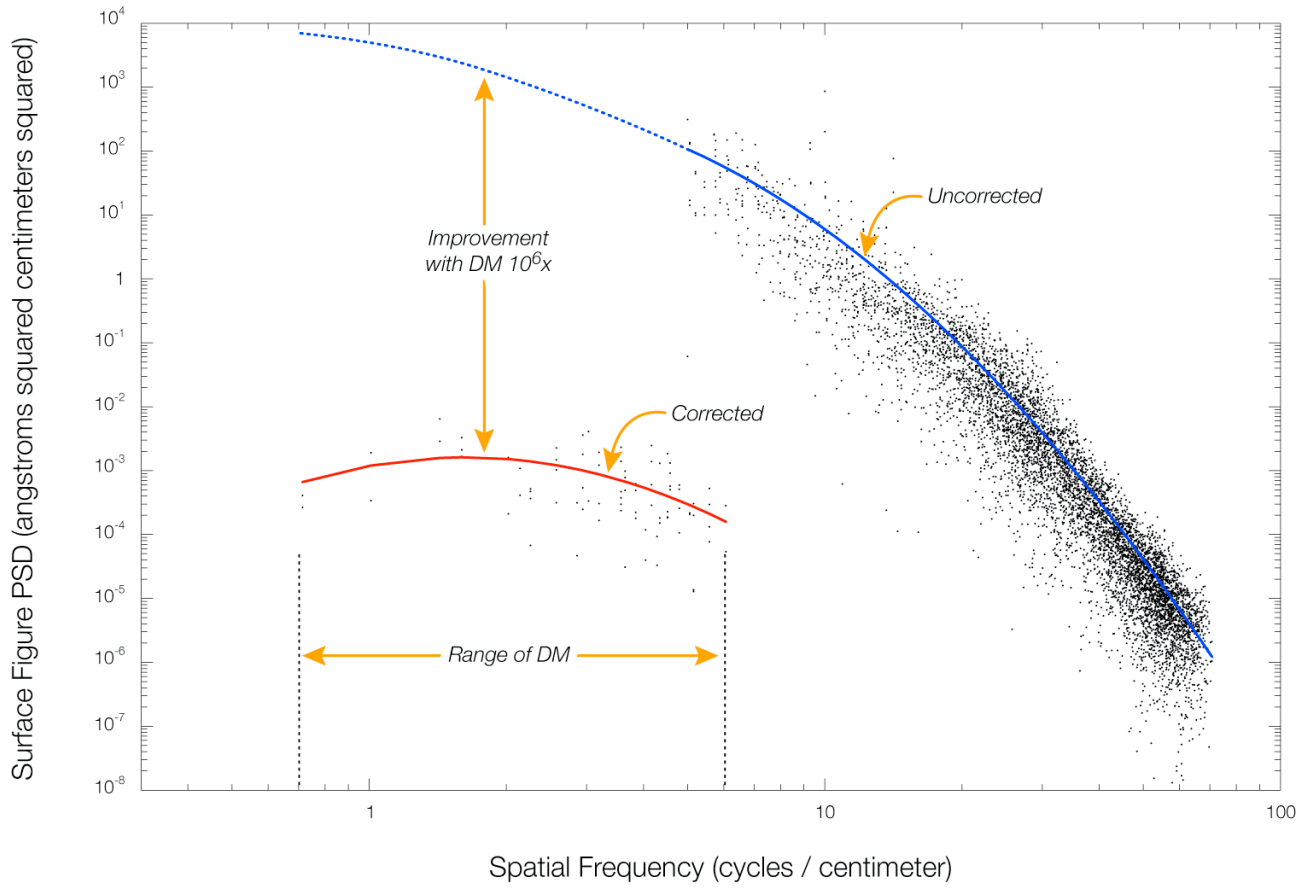

Fig. 1: Demonstration of deformable mirror (DM) technology with a surface metrology interferometer at JPL. The test results show that a 21 x 21 actuator DM can be commanded flat to 0.025 nm rms surface error (a wavefront error of 1/13000 waves at 632.8 nm) in the range of spatial frequencies under its authority, and the power spectrum of residual errors is flat (red line). The blue line shows the ripple errors at higher, uncorrected spatial frequencies.

3. THE 4mTPF

The 4mTPF is a 4-meter diameter, optical-ultraviolet space telescope with a coronagraph for deep suppression of starlight in an annular detection zone near a star. At 500 nm wavelength and with a deformable mirror having 200 x 200 actuators, the detection zone extends from 0.1 to 2.5 arcsec in radius around the star. The purpose of the deformable mirror is to correct wavefront errors. The coronagraph suppresses the surface brightness of diffracted starlight to (or below) 5 x10^{-11} times the peak value of the Airy diffraction pattern, which is the level of the speckle pattern of the PSDPSF for a DM with 1/7000 wave rms residual wavefront error. (For comparison, the average intensity of the theoretical Airy pattern of an unobscured 4-meter telescope at 500 nm and 0.1 (or 2.5) arcsec is already 7 x 10^{-4} (or 5 x 10^{-8}) of the peak value.)

A variety of coronagraph designs suitable for the 4mTPF have been proposed, including binary pupil-plane masks[7] and graded image-plane masks[8]. A hybrid design may emerge. The final choice of design awaits tradeoffs, such as between mask manufacturability (binary masks are simpler to make) and available search area (graded masks can allow greater azimuthal coverage).

The 4mTPF includes both low-resolution (resolving power R ~ 5) and medium-resolution (R ~ 70) photometric modes for color and biomarker studies in the spectral range 300 to 1100 nm.

The coronagraph uses only a small portion of the 4mTPF focal plane. As with the Hubble Space Telescope, other instruments, like a wider-field camera and a spectrograph, could be installed on the focal plane of 4mTPF to enable a general astronomy program.

3.1 Planet detection with the 4mTPF

We call a planet 'detectable' if the following two conditions hold: (1) At the detector, the central surface brightness of the planet image divided by the surface brightness of the local background—the contrast, Q—must be greater than or equal to unity, and (2) the observing system must be stable for the duration of the integration time needed to achieve a signal-to-noise ratio (SNR) of about five[9]. In the following, we translate the first condition into astronomical terms, estimate integration times, and discuss the stability aspects of the second condition as well as technical feasibility.

The surface brightness of an image on the detector is the flux incident on the telescope times the efficiency times the PSF. The 4mTPF coronagraph provides two point-spread functions: PSDPSF for the suppressed star when it is positioned at the center of the instrument field and PSF_{dz} for sources in the detection zone. The fluxes of the star and planet are S and $P(\alpha)$, where α is the planetocentric phase angle between star and observer. If $PSF_{dz}(0)$ is the central PSF value for planets and PSDPSF is the 'wing' or 'halo' value for the suppressed star at the planet position, the first detectability condition is:

$$Q \equiv P(\alpha)\, PSF_{dz}(0) / S\, PSDPSF \geq 1. \quad (1)$$

This neglects zodiacal light, which makes only a minor contribution at Solar-System levels for visible-light observations with small pixels, as with the 4mTPF. It also neglects the time dependence of PSDPSF, which is treated below.

The planet-to-star flux ratio (C) depends on planet radius (r), orbit radius (a), geometric albedo (p), and phase function (ϕ):

$$C \equiv P(\alpha)/S = p\, \phi(\alpha)\, (r/a)^2. \quad (2)$$

To illustrate using the case of Earth, we select λ = 500 nm and α = 90° and adopt p_{Earth} = 0.33 (taken from Fig. 3), $\phi_{Earth}(90°)$ = 0.29 (midway between the Lane and Irvine[10] value of 0.25 for a 'cloudy' planet and the value for a Lambertian sphere, $1/\pi$), and $(r_{Earth}/a_{Earth})^2 = 1.8 \times 10^{-9}$. In this case, $C = 1.7 \times 10^{-10}$.

We expect low-order aberrations and the coronagraph's pupil mask to reduce the planet image peak by a factor ~ 0.5 with respect to the theoretical Airy pattern. In this case the functional requirement on Airy suppression becomes $PSDPSF / PSF_{dz}(0) \leq 10^{-10}$, and Eqs. 1 and 2 indicate that detectability means $C > 10^{-10}$, or in terms of planet properties,

$$p\, \phi(\alpha)\, (r/a)^2 \geq 10^{-10}\ . \quad (3)$$

For Earth-like planets, which have Earth's photometric properties and its density, ρ_{Earth} = 5.5 gm cm^{-3}, the minimum detectable mass is

$$M_{planet} \geq 0.436\, a^3, \quad (4)$$

where M_{planet} and a are measured in Earth mass and AU. For planets in circular orbits with different density, geometric albedo, phase function, or phase angle, the minimum detectable mass is:

$$M_{planet} \geq 0.436\, [\rho/\rho_{Earth}]\, [\phi\, (\alpha)/\phi_{Earth}(90°)]^{-3/2}\, [p/p_{Earth}]^{-3/2}\, a^3. \quad (5)$$

While this result is independent of stellar distance, the inner and outer working distances in AU are not. At wavelength λ, the inner working distance is $4\lambda/D$ in radians, and outer working distance is $N\lambda/D$, where N ~ 100 is the number of ripple waves per aperture diameter D that are corrected by the ~ 2N deformable mirror actuators spaced across the re-imaged telescope pupil.

The discovery space for planets detectable according to the first (contrast) condition by the 4mTPF is shown in Fig. 2 for circular orbits and 90° phase angle. It also shows the 'habitable zone' for the Sun at 5 Gyr age in the most optimistic case of Kasting, Whitmire, & Reynolds[11]. (The habitable zone is the range of semi-major axis around the star in which water may occur in liquid form at the surface to support life.) Fig. 2 shows that 4mTPF is a true terrestrial planet finder in the habitable zones of nearby stars.

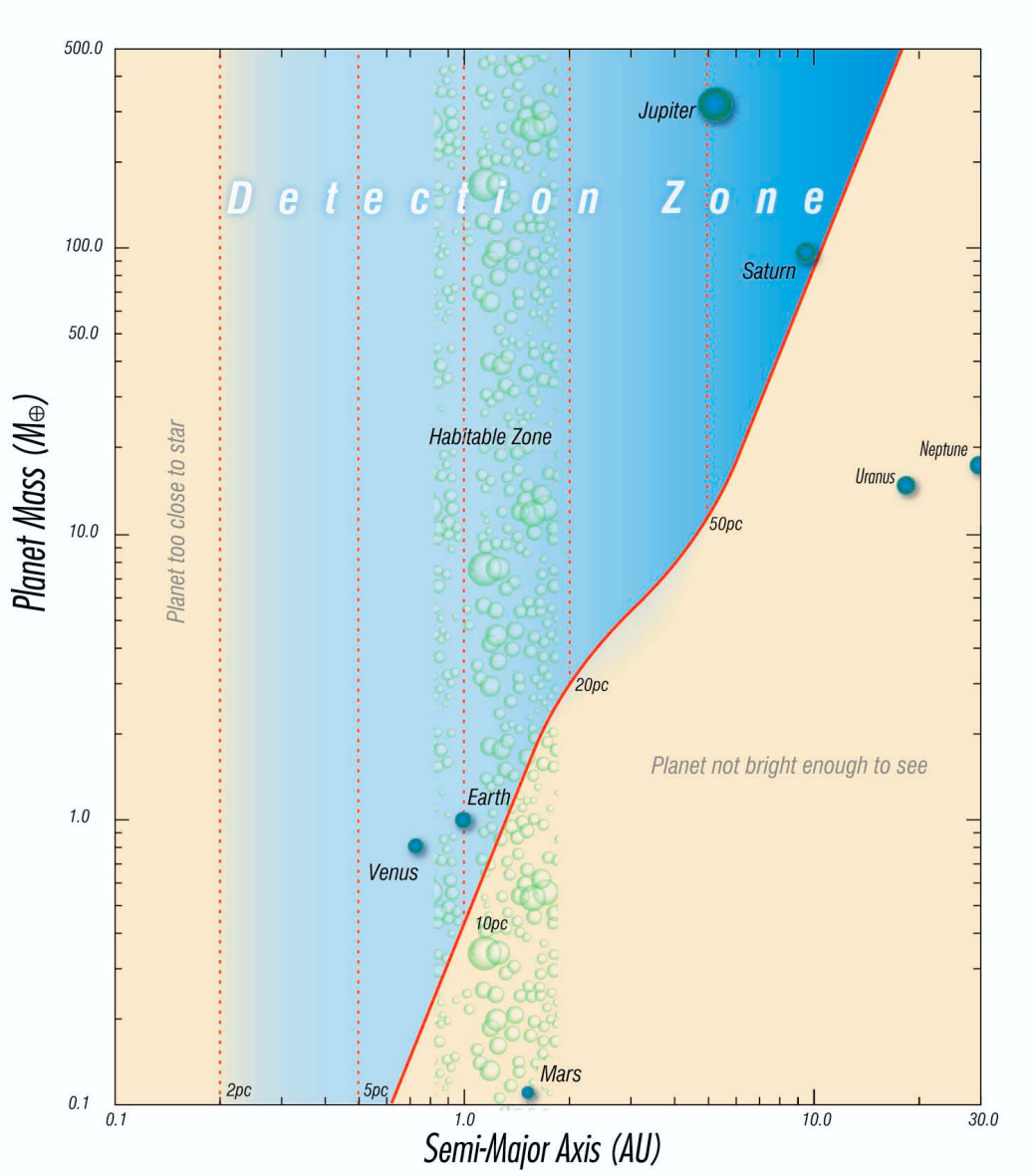

Fig. 2: The 4mTPF discovery space for extrasolar planets at λ = 500 nm and 90° phase angle, and in circular orbits. The diagonal red boundary line, independent of stellar distance, is drawn for an Earth-like density at lower mass and giant-planet densities at higher mass. The vertical dashed lines, which are located at 0.1 arcsec projected to various stellar distances, give the left-hand boundaries (inner working distance) of the 4mTPF discovery space. (The projected outer working distances (not shown) occur a factor of 25 to the right of the corresponding inner working distances.) The droplets indicate the habitable zone, where water can occur in liquid form. The 4mTPF robustly finds and studies Earth-like planets in habitable zones around Sun-type stars to 20 pc as well as Jupiter-like planets to beyond 50 pc.

3.2 Integration times

The integration time needed to detect a planet with a certain SNR depends on the planet flux and the total background surface brightness, which has two components: a fixed part that does not change between exposures and a variable part, which does change but has zero mean:

$$B_{tot} = B_{fix} + B_{var}, \text{ where } \langle B_{var} \rangle = 0. \quad (6)$$

B_{fix} contributes only photon noise to the photometric error, while B_{var} also contributes a systematic term.

In general, B_{fix} contains a contribution from residual starlight and smaller contributions from zodiacal light and extrasolar zodiacal light (if the extrasolar planetary system is like ours.)

Defining the two corresponding components of the contrast by

$$Q_{var} \equiv PSF_{dz}(0)\, P(\alpha)/\mathrm{sqrt}(\langle B_{var}^2 \rangle) \quad (7)$$

and

$$Q_{fix} \equiv PSF_{dz}(0)\, P(\alpha)/B_{fix}, \quad (8)$$

the total Q-factor is given by $Q = (1/Q_{fix}+1/Q_{var})^{-1}$, which we have required to be greater than or equal to unity.

If no other noise is present and a small number (n) of exposures of the planet and of a reference PSF (probably obtained by rotating the detection zone around the target star) are taken, then the planet peak surface brightness—and hence total planet flux—can be measured with a proportional error of roughly $1/Q_{var}$ sqrt(n). For example, SNR = 3 would require roughly nine exposures if Q_{var} were even as small as one.

Of course, we expect noise from a variety of sources. Of greatest importance is the photon noise from the fixed part of the stellar background, which contributes through Q_{fix}. The SNR for planet photometry is obtained by adding the Q_{fix} photon noise and Q_{var} systematic noise contributions in quadrature. If Q_{fix} is as small as one, the problem is equivalent to the problem of estimating the photon noise for photometry of a point source on a fixed and dominant background:

$$SNR^2 \equiv I^2\, \Psi/B, \quad (9)$$

where I is the number of photons detected from the source, B is the number of background photons per pixel, and Ψ is the sharpness,

$$\Psi = \Sigma(PSF^2)/(\Sigma PSF)^2. \quad (10)$$

($1/\Psi$ is the number of 'noise pixels.' For a critically sampled image from a perfect circular aperture, $\Psi \sim 0.07$.) If we assume reasonable values for the overall efficiencies, a 20% bandpass ($R_{det} \sim 5$) at λ_{det} =500 nm, $\Psi \sim 0.035$, then an Earth-equivalent planet at 5 pc is detected with $SNR_{det} \approx 5$ in an integration time $T_{det} \approx 1$ hr, where the subscript indicates detection rather than spectroscopy, which we consider next.

The integration time for spectroscopic detection of a feature (T_{spec}) is proportional to the resolving power matched to the feature's width (R_{spec}), the inverse square of the fractional depth of the feature (z) when observed at R_{spec}, the inverse of the stellar flux in the local continuum ($S(\lambda_{spec})$), the local albedo ($p(\lambda_{spec})$), the square of the desired SNR_{spec}, and a 'continuum factor' of order unity indicating how well the reference continuum is determined (=1 for known perfectly; ≈ 2 for continuum determined only at a few points near the feature with the same SNR_{spec}.) Assuming constant instrument performance and unit continuum factor, we can scale the result for T_{det}:

$$T_{spec} = T_{det}\, [SNR_{det}/ SNR_{spec}]^{-2}\, [R_{det}/ R_{spec}]^{-1}\, [1/z^2]\, [S(\lambda_{spec})/S(\lambda_{det})]^{-1}\, [p(\lambda_{spec})/p(\lambda_{det})]^{-1} \quad (11)$$

Using the A-band of O_2 at λ_{spec} = 760 nm, for which $R_{spec} \approx 69$ and $z \approx 0.47$ for a clear Earth-like atmosphere. Using the solar value $[S(500)/S(760)] \approx 0.61$ and assuming constant albedo, we estimate the integration time for detecting O_2 on an Earth at 5 pc with $SNR_{spec} \approx 5$ is $T_{spec} \approx 38$ hours.

3.3 Stability aspects

Our first detectability requirement is that Q be greater than one. In fact, although Q_{var} must closely satisfy this requirement, Q_{fix} can in principle be less than one. For example, infrared astronomers often perform observations with a fixed thermal background several orders of magnitude brighter than the objects they observe. In practice, however, several effects conspire to make 4mTPF-type observations with Q_{fix} much less than one very difficult. For example, for Q_{fix} = 1/100 (implying stable suppression at the 10^{-8} level), the integration time would be about 100 hours. While this might seem practical from a simple SNR perspective, our experience is that subtracting a spatially-variable PSF at the 1% level is extremely challenging and exposed to systematic errors, such as pointing fluctuations, detector variations, cosmic ray effects, and variations in charge transfer efficiency. We do not regard a 4mTPF-type experiment that is designed to operate at such low contrast levels as robust. Higher-resolution spectroscopy would require even more time and would quickly become impractical even from a simple SNR point of view. The stability requirement remains that Q_{var} be greater than 1, and in this case Q_{fix} can often be increased into the same range by suitable optical adjustments.

We have shown that if Q > 1, the full search space illustrated in Fig. 2 is accessible to imaging and medium-resolution spectroscopy in reasonable integration times—hours and tens of hours.

3.4 Technical feasibility

Existing technologies suffice to build and launch at least a 4-meter telescope with Hubble-like optical performance, and larger monolithic optics, particularly elliptical shapes, are feasible in the foreseeable future. Existing detector technology is adequate, too. Advanced technology is needed to control the phase, amplitude, and polarization of light in a stable manner. Phase-control technologies are being demonstrated in the laboratory now. Research and development is needed for amplitude and polarization controls. Stability issues can be addressed with both old technology (baffles, heaters, and thermal and mechanical isolation) and new technology (composite materials and active structures.)

To correct wavefront errors, the system design must include some combination of passive tolerancing and active control of the system wavefront in phase and amplitude. These tolerances and controls must apply to all polarization states allowed through the system. The simplest implementation—and the one discussed in this paper—involves a deformable mirror to control wavefront phase and passive tolerancing to set wavefront amplitude. With direct feedback from the final image plane (with no separate wavefront sensor and its associated calibration concerns), this scheme will correct at least half of the field over some finite bandpass. More complex schemes with wavefront and amplitude controls are available. For example, an interferometer with minor phase adjustments in one arm can control the amplitude effectively. If the wavefront amplitude is passively toleranced to a sufficient degree, then the full power of the deformable mirror can be used to correct the full field of view, not just half.

To achieve the detectability criterion Q > 1, any correction scheme requires that the system be stable, and this stability requirement is the fundamental technical uncertainty and design driver. The stability constraint on the rms variations in throughput is $\sigma_t < \pi\ N\ s^{1/2}\ C^{1/2} / 2$, and that on phase is $(\sigma_p / \lambda) < N\ s^{1/2}\ C^{1/2} / 4$. (For 4mTPF just detecting Earth, $\sigma_t \sim 0.2$ % and $\sigma_p \sim .13$ nm or 1/5000 wave.) We expect the variations to be uncorrelated if they arise randomly due to, for example, micrometeorite events or errors in the placement of actuators in an active control system.

The specifications of amplitude and phase transmission stability are exacting, for the optics are susceptible to mechanical vibrations and thermal distortions at the critical spatial frequencies. Modelling and testing these effects are consuming much current effort.

We believe that the element in the optical train of greatest concern is the primary mirror. Preliminary results of our study show that using existing components to isolate the structure from sources of disturbance (e.g. reaction wheels) can adequately suppress mechanical vibrations. (One caveat is that low-level structural effects such as micro-snap and creep are currently not well understood and will require further analysis and testing.) The highest technical risk is that of

wavefront changes that occur too rapidly to measure with the finite number of scattered photons. The primary mirror must have low, uniform thermal expansion coefficient and low internal stresses. A suitable orbit and baffle design must shield the primary mirror from radiation from the Sun, Earth, and Moon.

4. PLANETARY SCIENCE WITH THE 4M

The 4mTPF can observe any planet brighter than the limiting case between the inner and outer working distances. Its most pertinent targets are Earth-like planets around nearby solar-type stars, which the 4mTPF can discover and study to determine their properties and to estimate their habitability. The 4mTPF can also observe extrasolar giant planets, which can influence conditions on Earth-like planets.

The search sample for the 4mTPF consists of nearby solar-type, single stars. Operating at $\lambda = 500$ nm with an inner working distance of 0.1 arcsec, the 4mTPF can explore the entire habitable zones of 20 stars of spectral types A0 to K0, mostly late F- to G-type. For 150 A0 to K8 stars, the 4mTPF can search at least part of the habitable zone. There are 32 stars for which > 95% of the habitable zone volume is visible, 29 stars with 80 - 95%, 32 with 65 - 80%, and 18 with 50 - 65%. In total, there are 111 stars for which > 50% of the volume of the habitable zone is accessible.

The observatory-level science capabilities of the 4mTPF are its abilities to (1) discover planets, (2) measure their spectral flux at low and medium resolving powers, (3) measure their apparent locations—separations and position angles—relative to their stars, and (4) make these measurements at various epochs to study variations due to orbital motion, changing phase-angle, and rotation, as well as to measure intrinsic variability, such as might be produced by climate or weather.

The 4mTPF does not resolve the planet; it integrates the light from the entire visible portion of the illuminated planetary disk. Each photon has undergone a complex interaction with the planet, including two transits through any atmosphere and perhaps multiple scatterings by gases, liquids, or solids. Despite these complications, but in part due to these interactions, the 4mTPF gathers a wealth of information.

4.1 Mass determination

Planetary mass provides clues about a planet's formation history and its ability to retain an atmosphere and foster life. Each of these applications requires a sufficiently accurate value of mass to draw pertinent distinctions. For example, an order-of-magnitude in mass separates the atmosphere-retaining and plate-tectonic-capable Earth and Venus from Mars, which has a minor atmosphere and no plate tectonics. Another order of magnitude distinguishes Earth and Venus from Uranus and Neptune, which have thick atmospheres. Jupiter and Saturn, with overwhelmingly thick atmospheric layers, are yet another order-of-magnitude step in mass. Thus, mass accuracy of a factor 2 is adequate to classify extrasolar planets in the Solar System taxonomy, which is a valid initial framework.

To measure the mass of a planet, the 4mTPF obtains images in two or more low-resolution photometric passbands, time-spaced to show orbital motion for estimating the planet-star distances and phase angles in the images. The observed phase function helps distinguish between a rocky, cloudy, or Rayleigh-scattering atmosphere planet. We compare the wavelength variation of flux with the four-color types of Solar System planets, shown in Fig. 3.

The first color type—Mercury, Mars, and the Moon—comprises low-mass, rocky planets with very low albedo in the blue, increasing smoothly toward the red. No atmospheric signature is present. Since the time scale for atmospheric mass loss is quite short, the lack of an atmosphere implies low mass. The second color type—Jupiter and Uranus—are gas-giant planets with clouds and Rayleigh-scattering atmospheres with high reflectance in the blue and green. Large amounts of methane and ammonia produce a precipitous drop in albedo in the red. The thick atmosphere of light, primitive gases implies high mass. Venus represents the third color type: dark in the blue and bright in the red due to strong scattering from cloud particles with some ultraviolet-absorbers. Earth is the fourth color type: strong blue albedo due to Rayleigh scattering in a relatively clear atmosphere and flat green and red albedo from reflection off clouds; a 'red-edge' enhanced albedo in the red may be seen if land plant leaves are abundant. For both the third and fourth color

types, the lack of red absorption due to primitive gases implies that the mass is intermediate between a gas-giant and a rocky planet.

Based on our Solar-System knowledge and many independent clues to the physical characteristics of a planet—flux, color type, phase curve—we can constrain its mass to within a factor of 2.

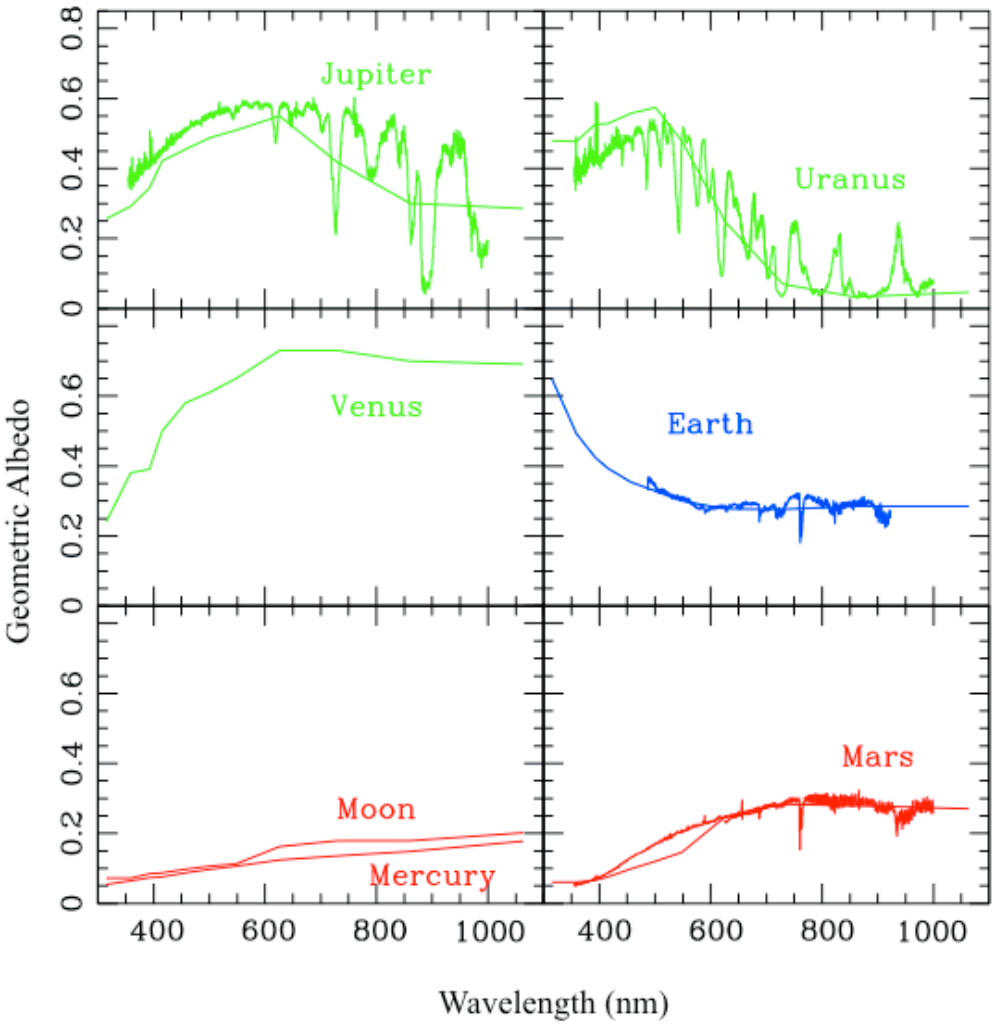

Fig. 3: Geometric albedos of six Solar System planets and the Moon. The smooth curves are integrated disk photometric data[12], except for the Earth, where a flat spectrum plus a Rayleigh and red-edge component are inferred from the spectral data from Ref. 20. The other spectra are recent data scaled to match the photometric data in a central wavelength band (W.A.T., Berlind, P., and Jucks, K. W., in preparation.) The higher-resolution spectra illustrate why the low-resolution ('color') photometry varies with wavelength as it does. For an extrasolar planet, low-resolution photometry (R ~ 5) would suffice to distinguish between these planet types.

4.2 Orbit determination

Orbital parameters are essential for understanding the thermophysical conditions on a planet, the dynamical structure of a system of planets, and the processes involved in planet formation. For example, coplanar, low-eccentricity orbits are clues that planet formation took place in a flattened, viscous disk around the young star, as is thought to be true for the Solar System. An accuracy of 10% is adequate when determining the period, semi-major axis, eccentricity, and inclination for such purposes.

The 4mTPF determines the orbit by measuring the apparent position of the planet as it revolves around the star. Given ideal optics, attitude, and detectors, the accuracy of the object's location within the image—and therefore of the separation and position angle on the sky—is limited only by photon statistics. In this case, the expected positional uncertainty for a point-source image is:

$$\sigma_{x \text{ or } y} = K \lambda/(4\pi d \, SNR), \tag{12}$$

where K is a degradation factor of order 1 that takes into account geometrical aberrations, pixel-sampling, the position-measuring procedure, and various sources of noise[13,14]. For the ideal case, K = 1; more realistically K ≥ 2. Any real system will impose systematic limits on this equation for high SNR, due to instabilities, calibrations, manufacturing tolerances, and other sources of systematic error.

For the conservative case of 4mTPF imagery at λ = 500 nm, SNR = 5, and K = 2, we find $\sigma_{x \text{ or } y}$ = 0.8 millarcsec or 250 times smaller than the 0.2 arcsec the semi-major axis of an Earth-equivalent planet subtends at 5 pc. We expect the position of the star image with respect to the detector is fixed and stable to higher accuracy. In the absence of a comprehensive Monte Carlo study of the accuracy with which 4mTPF images determine orbital parameters, we can form expectations from related experience. For example, the Fine Guidance Sensors on the Hubble Space Telescope measured the M-dwarf binary Wolf 1062 using differential astrometry[15]. The 14 measures, well distributed about the 2.4-year orbit yet not quite providing 100% coverage, have a precision of 1 milliarcsec. Judging from the orbital residuals, the average random error is 1.4 milliarcsec on each coordinate. The typical measured separation is about 100 times larger. The formal errors in the resulting period, semi-major axis, eccentricity, and inclination angle are less than about 1%. These finding are in agreement with similar analyses in the literature[16,17].

For the 4mTPF, we expect positional uncertainties to be less than 10% of typical measured separations, which implies that the 4mTPF can measure the period, semi-major axis, eccentricity, and inclination of extrasolar planets to better than 10%.

4.3 Biomarkers in the visible spectrum

The visible spectrum of an extrasolar Earth-like planet contains a rich set of up to seven biomarker features, which can tell us much about the planet and even if life is present. The most important of these are O_2, O_3, H_2O, and the 'red-edge' feature of chlorophyll-containing plants, followed by the total molecular column, CO_2, and CH_4[18,19,20]. Table 1 provides quantitative information about these biomarkers, including the estimated integration time for the 4mTPF to detect them with SNR_{spec} = 5 on an Earth at 5 pc, based on Eq. 11.

For species relevant to the present Earth, the projected integration times are all less than 2 days. For example, the Rayleigh-scattered blue feature is detected in 1 hour, H_2O in 4 hours, O_3 in about 30 hours, O_2 in about 40 hours, and the red edge of plant life in 40 hours. If we encounter an early-Earth-type planet, we might expect to see CH_4 in about 6 hours and CO_2 only after a rather long 2 months.

Oxygen (O_2) is deservedly a widely-recognized sign of life. We believe that only large masses of plant life can create significant amounts of O_2. Photochemical processes—starting with the photolysis of H_2O or CO_2 by short-wavelength sunlight—easily explain the small amounts of O_2 in environments like Mars, Venus, and Europa. Because oxygen provides a high-grade fuel for animals to metabolise, it indirectly indicates the possibility of animal-type life forms. The 760 nm band of O_2, the Fraunhofer A-band, provides the 4mTPF with a strong indicator of O_2.

Ozone (O_3) is a by-product of O_2, created by photolysis of O_2 and combination to form O_3. Even small amounts of oxygen can diffuse into the upper atmosphere of a planet where the star's ultraviolet light can convert it to relatively long-lived ozone. Because of its strong spectral features, O_3 is a robust, nonlinear indicator of the presence of O_2, but it is a poor measure of the amount of O_2. Based on our understanding of the young Earth, early O_2 production would lead to a relatively strong O_3 stratospheric feature even if the surface were a sink for O_2. In the visible, the broad Chappuis band in the middle of the green part of the spectrum provides the 4mTPF a sure indicator of ozone. If the ultraviolet is also accessible, then the extremely strong O_3 band below about 300 nm is another excellent indicator.

Feature	λ	T_{spec}	Flux	R_{spec}	z	Abundance
O_2	760	38	1.63	69	0.47	Present (21% uniformly mixed)
O_3	580	27	1.26	5	0.17	Present (6 ppm peak in stratosphere)
O_3	320	33	0.23	16	0.65	Present (6 ppm peak in stratosphere)
H_2O	940	4	1.67	17	0.76	Present (8000 ppm at surface = 50% saturation)
H_2O	720	28	1.58	37	0.41	Present (8000 ppm at surface = 50% saturation)
red edge	800	42	1.66	5	0.12	Present (emission feature, Earthshine)
Rayleigh	400	1	0.56	5	1.10	Present (emission feature, Earthshine)
CO_2	1050	1366	1.63	40	0.06	Early Earth (10% = 0.1 bar)
CH_4	890	6	1.68	32	0.82	Early Earth (1% = 0.01 bar, methane burst)
CH_4	730	34	1.59	57	0.46	Early Earth (1% = 0.01 bar, methane burst)

Table 1: Integration times for detecting biomarkers for Earth/Sun at 5 pc. 'Feature' is the name of species or other name for the biomarker. 'λ' is the wavelength of the spectral feature in nm. 'T_{spec}' is the integration time in hours to achieve SNR_{spec} = 5 per Eq. 11, assuming constant albedo and T_{det} = 1 hour for SNR_{det} = 5, R_{det} = 5, and λ = 500 nm. 'Flux' is the solar flux at λ divided by the solar flux at λ = 500 nm. 'R_{spec}' is the resolving power that matches the width of the spectral feature. 'z' is the depth of the spectral feature. 'Abundance' describes the amount of the gas and the associated scenario. The strengths of the red edge and Rayleigh features are scaled from Earthshine observations presented in Fig. 2 of Ref. 20.

Liquid water (H_2O) is required for life forms familiar to us, and we expect that other forms of life would require it as well. This expectation is contained, for example, in the definition of the habitable zone around a star. The habitable zone can actually be significantly wider than might be expected from the albedo of a planet, because the greenhouse effect can make the surface warmer than the planet's effective temperature might indicate. In the case of Earth, for example, the effective radiating level is high in the atmosphere, where the local temperature is well below freezing.

Chlorophyll-containing green-leaf plants on land reflect strongly in the range 720-1000 nm. Earth-oriented remote-sensing satellites routinely use this 'red-edge' feature to sense the presence of plant life. The feature has even been tentatively identified in the integrated-light spectrum of the Earth ('Earthshine') as seen in light reflected off the dark side of the Moon[20,21]. A 4mTPF discovery of the red-edge feature would be a highly suggestive sign of life, because there is no known substance that mimics this feature.

Rayleigh scattering reveals the molecular column or total surface pressure of atmospheric gas in a relatively clear atmosphere. On Earth, it makes the famous 'pale blue dot,' which refers more to the blue of the sky than the rather dark blue of the ocean. The 'pale' part is due to the dilution of this effect by reflected sunlight from bright cloud tops. In fact, cloud reflection and Rayleigh scattering make up most of the optical-ultraviolet flux of the Earth. No other Solar System planet exhibits the blue reflection spectrum of the Earth.

Carbon dioxide (CO_2) was abundant on the early Earth, and strong enough to detect in the visible spectrum extended to about 1100 nm. Today's relatively small amount of CO_2 on Earth creates a much smaller absorption feature, which would not have diagnostic value. Nevertheless, since the early Earth is a possible model for an extrasolar planet, the 4mTPF might include this wavelength region in its spectroscopic capability.

It is thought that there were several 'methane burst' events in the Earth's history, sufficient to cause an easily measurable feature. The source of methane might have been the undersea methane clathrate material that we see today on the ocean floor in great abundance. A disturbance, such as a change of pressure or temperature, would be sufficient to destabilize this material and convert it to gaseous form.

4.4 Photometric variations.
The 4mTPF obtains diurnal light curves of extrasolar planets, which contain information about rotation periods, major surface features, and meteorological activity[22]. As they rotate in and out of view, continents, oceans, ice/snow sheets, and gradually changing large-scale cloud systems could produce measurable (tens of percent) changes in the flux and color of the integrated scattered starlight from a planet. Because half or more of the light typically comes from 10% of the planet's surface in a typical viewing geometry, a light curve can 'resolve' an extrasolar planet at that level.

The 4mTPF will also record photometric variations with phase angle as a planet revolves in its orbit around the star. Variations on an orbital timescale may indicate seasonal climate variations or the presence of planetary rings[23].

5. FINAL REMARKS

The 4mTPF could include other astronomical instruments on the focal plane to enable a general program of astronomical research like that of the Hubble Space Telescope. Currently, NASA plans to terminate Hubble operations in 2010 and has no plan to replace its unsurpassed imaging and spectroscopic capabilities at optical-ultraviolet wavelengths. (The Next Generation Space Telescope (NGST) is a cooled, infrared telescope, which will complement—but not replace or extend—Hubble's unique observing power at wavelengths below 600 nm.) The focal plane of the 4mTPF—as well as its mission—can accommodate other instruments, like a wider-field camera and a spectrograph, without interfering with the function or operation of the coronagraph.

The 4mTPF sets a fiducial against which we may measure other systems and techniques proposed to find and study Earth-like planets around nearby stars. We are confident that such comparisons will demonstrate the merit and priority of the 4mTPF on the basis of performance (SNR, integration time) and science program.

A single optical-ultraviolet space telescope could be both a Terrestrial Planet Finder, and a worthy successor to Hubble Space Telescope.

ACKNOWLEDGMENTS

We thank Richard Goody and Jean Schneider for their critical reviews and helpful suggestions for improving our paper. We thank Otto Franz for his discussion of orbit determination.

REFERENCES

1. NASA. AO No. OSS-1-77: "Announcement of Opportunity for Space Telescope." (1977).
2. Elson, J. M., Rahn, J. P. & Bennett, J. M. "Relationship of the total integrated scattered light from multilayer-coated optics to angle of incidence, polarization, correlation length, and roughness cross-correlation properties." *Appl. Opt.* **22**, 3207-3219 (1983).
3. Elson, J. M. "Theory of light scattering from a rough surface with an inhomogeneous dielectric permittivity." *Phys. Rev. B* **30**, 5460-5480 (1984).
4. Brown, R. A. & Burrows, C. J. "On the feasibility of direct extrasolar planet detection using Hubble Space Telescope." *Icarus* **87,** 484-497 (1990).
5. The Space Telescope Science Institute and Ball Aerospace Systems Division prepared and submitted the CODEX proposal in response NASA's announcement of opportunity AO 96-OSS-03. NASA rated the proposal "selectable" but has not yet selected it for flight.
6. J. T. Trauger, et al., "Performance of a precision high-density deformable mirror for extremely high contrast imaging from space," SPIE paper 4854-01.
7. D. S. Spergel, "A new pupil for detecting extrasolar planets," http://xxx.lanl.gov/abs/astro-ph/0101142, 2002.
8. M. J. Kuchner and W. A. Traub, "A coronagraph with a band-limited mask for finding terrestrial planets," *Astrophys. J.* **570,** 900-908, 2002.
9. R. A. Brown, "Systematic Aspects of direct extrasolar planet detection using Hubble Space Telescope," in *Bioastronomy: The Next Steps.* (G. Marx, ed.) Dordrecht: Kluwer Academic Press., 1988.
10. A. P. Lane and W. M. Irvine, "Monochromatic phase curves and albedos for the lunar disk," *Astrophys. J.* **78**, 267-277, 1973.
11. J. F. Kasting, D. P. Whitmire, and R. T. Reynolds, "Habitable zones around main sequence stars," *Icarus* **101,** 108-128, 1993.
12. W. M. Irvine, T. Simon, D. H. Menzel, C. Pikoos, and A. T. Young, "Multicolor photoelectric photometry of the brighter planets. III. Observations from Boyden Observatory" *Astron. J.* **73,** 807-828, 1968.

13. L. Lindegren, "Photoelectric astrometry—a comparison of methods for precise image location," in *Modern Astrometry: Proceedings of the IAU Colloquium No. 48*. (F. V. Prochazka & R. H. Tucker, eds.) Vienna, Institute of Astronomy, 1978.
14. M. Gai, D. Carollo, M. Delbo, M. G. Lattanzi, G. Massone, F. Bertinetto, G. Mana, and S. Cesare, "Location accuracy limitations for CCD cameras," *Astron. Astrophys.* **367,** 362-370, 2001.
15. O. G. Franz, *et al.*, "The first definitive binary orbit determined with the Hubble Space Telescope Fine Guidance Sensors: Wolf 1062 (Gliese 748)," *Astron. J.* **116,** 1432-1439, 1998.
16. A. Sozzetti, S. Casertano, M. G. Lattanzi, and A. Spagna, "Detection and measurement of planetary systems with GAIA," *Astron. Astrophys.* **373**:L21-L24, 2001.
17. A. Sozzetti, S. Casertano, R. A. Brown, and M. G. Lattanzi, "Narrow-angle astrometry with the Space Interferometry Mission: the search for extra-solar planets. I. Detection and characterization of single planets," P. A. S. P., in press. 2002.
18. D. J. Des Marais, M. Harwit, K. W. Jucks, J. Kasting, D. Lin, J. Lunine, J. Schneider, S. Seager, W. A. Traub, and N. J. Woolf, "Remote sensing of planetary properties and biosignatures on extrasolar terrestrial planets," *Astrobiology,* in press, 2002.
19. W. A. Traub and K. W. Jucks, "A possible aeronomy of extrasolar planets." In *Atmospheres in the Solar System: Comparative Aeronomy.* (M. Mendilo, A. Nagy, and H. J. Waite, eds.) *AGU Geophysical Monograph* **130,** 369-380, 2002.
20. N. J. Woolf, P. S. Smith, W. A. Traub, and K. W. Jucks, "The spectrum of Earthshine: a pale blue dot observed from the ground," *Astrophys. J.* **574**, in press, 2002.
21. L. Arnold, S. Gillet, O. Lardiere, P. Riaud, and J. A. Schneider, "A test for the search for life on extrasolar planets: Looking for the terrestrial vegetation signature in the Earthshine spectrum." *Astron. Astrophys.* in press, 2002.
22. E. B. Ford, S. Seager, and E. L. Turner, "Characterization of extrasolar terrestrial planets from diurnal photometric variability," *Nature* **412,** 885-887, 2002.
23. J. Schneider, "The study of extrasolar planets: Methods of detection, first discoveries, and future perspectives," *C. R. Acad. Sci. Paris* **327,** *Serie II b*, 621-634, 1999.